

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-219312

(43) 公開日 平成11年(1999) 6月10日

(51) Int.Cl.⁴
G 0 6 F 12/00

識別記号

5 4 6

5 4 7

5/00

13/00

5 5 4

F I

G 0 6 F 12/00

5 4 6 K

5 4 7 D

5/00

H

13/00

5 5 4 Z

審査請求 未請求 前請求の件12 O L (全 11 頁)

(21) 出願番号 特開平10-22227

(22) 出願日 平成10年(1998) 2月4日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田墨河台西丁1番6番地

(72) 発明者 西原 真二

東京都墨田区市東区ケ盛一丁目280番地

株式会社日立製作所中央研究所内

(72) 発明者 西原 真

東京都墨田区市東区ケ盛一丁目280番地

株式会社日立製作所中央研究所内

(72) 発明者 牛嶋 一智

東京都墨田区市東区ケ盛一丁目280番地

株式会社日立製作所中央研究所内

(70) 代理人 弁理士 小川 勝男

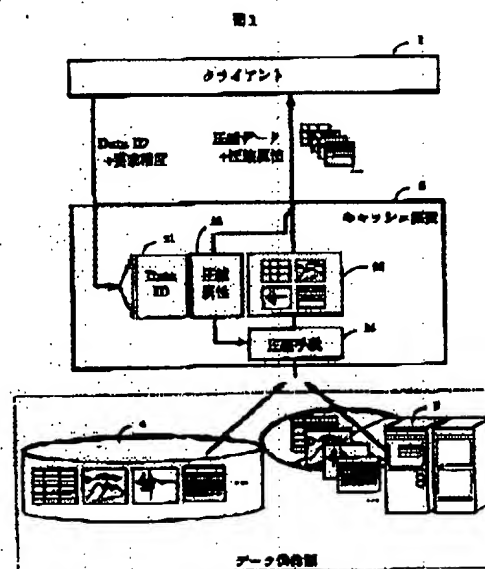
最終頁に続く

(54) 【発明の名称】 データキャッシュ方法およびデータアクセス方法

(57) 【要約】

【課題】サイズが大きいデータオブジェクトをデータの精度を指定して効率よくアクセスするためのデータキャッシュ方法およびデータアクセス方法を提供する。

【解決手段】クライアントからデータの要求精度を含むデータアクセス要求を発行する。キャッシュ装置は指定されたデータがキャッシュ内に存在するかどうかを判定し、指定されたデータが存在した場合にはキャッシングされたデータの圧縮属性を参照し、データアクセス要求に指定されている精度を満たすならば、その圧縮データを返す。キャッシュに存在しないかあるいは精度が満たされなかった場合には、データ供給源から受け取ったデータを指定された精度を満たす範囲で圧縮してキャッシュに登録してクライアントに返す。



特開平11-219312

(2)

1

【特許請求の範囲】

【請求項1】データを一意に識別するための識別子配列とデータを格納するキャッシュ領域とを有するキャッシュ装置の、前記キャッシュ領域にはデータの精度を落とす不可逆圧縮方法が適用された圧縮データを格納し、上記データの圧縮に関する圧縮属性をデータとともにキャッシュ装置内に格納し、データをアクセスするクライアントはデータの精度を指定したデータアクセス要求を前記キャッシュ装置に対して発行し、キャッシュ装置はクライアントから要求されたデータがキャッシュ内に存在するかどうかを前記データ識別子配列を検索することで判定し、データが存在する場合には、上記データに対応する圧縮属性とデータの要求精度とを比較し、クライアントのデータ要求精度が満たされている場合には圧縮されたデータを圧縮属性とともにクライアントに返送することを特徴とするデータキャッシュ方法。

【請求項2】前記データキャッシュ装置に対するデータの精度を指定したアクセス要求が、上記キャッシュミスした場合に、クライアントから指定された精度を満たすデータ精度を指定したアクセス要求をデータ供給源または後続のキャッシュ装置に対して発行することを特徴とする請求項1記載のデータキャッシュ方法。

【請求項3】前記データキャッシュ装置に対するデータの精度を指定したアクセス要求がキャッシュミスした場合に後続のデータ供給源またはキャッシュ装置に対して発行したデータアクセス要求に対して返されたデータが圧縮されていないオリジナルのデータであった場合には、上記キャッシュ装置内において、クライアントから指定されたデータ精度を満たす範囲で圧縮を行い、上記圧縮データをクライアントに返送するとともに、前記キャッシュ装置内に圧縮属性とデータ識別子とともに登録することを特徴とする請求項1記載のデータキャッシュ方法。

【請求項4】クライアントからのデータアクセス要求にデータの精度の指定がなかった場合にはオリジナルデータを要求しているとして圧縮キャッシュをアクセスしないことを特徴とする請求項1記載のデータキャッシュ方法。

【請求項5】データを圧縮して格納するための圧縮データキャッシュに格納する圧縮データとして、データの精度を落とした不可逆圧縮方法を適用した圧縮データに加えて、データの精度を落とさない可逆圧縮方法を適用した圧縮データも格納し、上記可逆圧縮されたデータの精度はすべての精度付きアクセス要求を満たす扱いをすることを特徴とする請求項1記載のデータキャッシュ方法。

【請求項6】キャッシュ装置にデータを圧縮して格納するための圧縮データキャッシュに加えて、データを圧縮せずに格納する通常のデータキャッシュを備え、データをアクセスするクライアントがデータの精度を指定した

2

アクセス要求を発行し、上記データアクセス要求が、圧縮データキャッシュにヒットし、かつ、データ要求精度が満たされている場合には、通常のキャッシュに対するデータアクセス要求のヒットの有無に関わらず圧縮されたデータを優先的にクライアントに戻すことを特徴とするデータキャッシュ方法。

【請求項7】オリジナルのデータを保持するデータ供給源において、上記データ供給源は、オリジナルデータを圧縮した圧縮データを保持し、データの精度が指定されたアクセス要求に対して、前記圧縮によるデータの精度の低下がアクセス要求で指定された精度を満たしているかどうかを判定し、精度が満たされているときには圧縮されたデータを返送することを特徴とするデータアクセス方法。

【請求項8】オリジナルのデータを保持するデータ供給源において、上記データ供給源が圧縮されたデータを保持しないか、あるいは、上記データ供給源が保持する圧縮データの精度がアクセス要求で指定された精度を満たさなかったときに、上記データ供給源において、オリジナルデータをアクセス要求で指定された精度を満たすように圧縮して、上記圧縮データを返送することを特徴とするデータアクセス方法。

【請求項9】請求項3におけるオリジナルデータを圧縮して圧縮データを作成する処理において、クライアントからの精度付きのデータアクセス要求の履歴をキャッシュ装置内に保持することにより、クライアントからのデータの要求精度を満たす範囲で、かつ、過去の上記データに対するデータの要求精度を可能な限り満たす範囲で、データのキャッシュング精度を決定し、圧縮方法を決定することを特徴とする請求項3記載のデータキャッシュ方法。

【請求項10】オリジナルデータを圧縮した圧縮データを保持するキャッシュ装置において、前記圧縮方法はデータの精度を落とす不可逆圧縮が適用されており、上記キャッシュ装置に対してクライアントが発行するデータアクセス要求にデータの精度が指定されていることを特徴とするデータアクセス方法。

【請求項11】請求項1記載のデータの精度を落とす圧縮方法において、データベースのテーブルに対しては、サンプリングされたデータを保持することによりデータ圧縮を行うことを特徴とするデータ圧縮方法。

【請求項12】請求項1記載のデータの精度を落とす圧縮方法において、テキストデータに対しては、テキストデータの要約またはキーワードリストを保持することにより圧縮を行うことを特徴とするデータ圧縮方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、情報システムにおけるデータオブジェクトのキャッシュ方法ならびにデータアクセス方法に関する。

(3)

特開平11-219312

【0002】

【従来の技術】情報化社会の発展と計算機技術の進歩に伴い、テキスト情報だけでなく、音声、画像、動画など、各種のマルチメディア情報がシステム内でやりとりされるようになってきた。これらの情報は従来のテキスト情報とは異なり、1オブジェクトあたりのデータサイズが非常に大きいという特徴がある。また、テキスト情報でも、論文や雑誌の全文を一つのオブジェクトとするようなデータベースが構築されてきており、情報システム全体で扱うデータ量はきわめて膨大になってきた。さらに、各企業では膨大な生データを蓄積してデータ解析を行うデータウェアハウスシステムの導入が急速に進んできており、データベースシステムが保持するテーブルも大きなシステムでは数テラバイトに及ぶ量となってきている。

【0003】これらの膨大な情報をインターネット上で共有するためには、これらの情報を効率よくアクセスする技術が必須となる。とくに、インターネット上のWorld Wide Web (以下、WWW) のようなシステムでは、データ量の増大に伴い、ネットワークにかかる負荷が大きくなり、データ転送に要する時間が長くなるという問題があった。

【0004】このような問題を解決するための技術として、キャッシングサーバがある。キャッシングサーバは、インターネット上でのWWWシステムの発展に伴い出現してきた技術であり、例えば、「Michael Beentzen, et al., "World Wide Web Caching: The Application-Level View of the Internet", IEEE Communications Magazine, June 1997,」などにその歴史等が記されている。この中では、キャッシュの一直性を保つ技術やキャッシングサーバの負荷分散に関する技術が述べられている。

【0005】映像情報などの膨大なデータをキャッシングサーバに格納すると、他の有益なデータが追い出される可能性が高くなる。これを回避するために、キャッシングサーバに格納するデータのサイズを制御する仕組みとして、あるサイズ以上のデータはそのままキャッシングせず、データを圧縮してキャッシングする方法がある。この従来方法については、「M. Ouchi and K. Ono, "A Proposal for World-Wide Web Caching Proxy Mechanism", In Proc. of Third International Symposium on Interworking (INTERWORKING'96), pp. 531-540, October 1996,」に述べられている。

【0006】ここで述べる圧縮とは、デジタルまたはアナログデータを、本来のサイズより小さくするとき（圧縮）に用いられる方法のことである。圧縮には、データを伸張する際に、元データのデータに復元する歪みを許さない可逆圧縮 (Lossless) と、元のデータよりも情報量が不足する可能性がある歪みを許す不可逆圧縮 (Lossy) があり、文字情報やプログラムなどの圧縮には、一般

的に歪みを許さない圧縮法を用いてデータサイズを縮小し、動画画像や音声データなどは歪んでいい程度に情報の歪みを許す圧縮法が用いられる。

【0007】上記歪みを許さない可逆圧縮にはランレングス法やハフマン法などの圧縮アルゴリズムがある。また、画像（静止画/動画）圧縮に用いられるJPEGやMPEGは、歪みを許さない可逆圧縮と歪みを許す不可逆圧縮の双方を取り入れた圧縮方式であり、歪みを許す指定を行った場合には、特に高い圧縮率を得ることができ

る。

【0008】不可逆圧縮では、オリジナルデータを再現できないため、データの品質が劣化するが、映像や画像などのマルチメディアデータでは必ずしもオリジナルデータと同じ品質のデータをユーザが要求しているとは限らない。そこで、この論文では、クライアントからのアクセス要求に対して、圧縮されたキャッシュにヒットした場合には、クライアントにこの圧縮に基づく品質で良いかどうかを問い合わせ、クライアント側がその品質を許容した場合にはその圧縮データを転送するプロトコルが明示されている。

【0009】図8はこの論文で述べられている圧縮キャッシュを利用するプロトコルを説明する図である。

【0010】図8(a)はクライアントの要求がキャッシュ装置（キャッシュサーバ）の圧縮キャッシュにヒットし、クライアントが当該圧縮によるデータの品質の劣化を許容した場合の処理の流れを示す図である。まず、クライアント1がキャッシュ装置2に対してデータアクセスを要求する(801)。次に、キャッシュ装置は、装置内に保持するオリジナルデータのキャッシュ81と圧縮データのキャッシュ82を調べ、該当するデータがあるかどうかをチェックする。ここで、オリジナルデータのキャッシュにはデータがなく（キャッシュミス）、圧縮キャッシュにのみデータがあった（キャッシュヒット）場合には、キャッシュ装置はクライアントに対して、圧縮キャッシュ内に格納されているデータの精度を応答する(802)。クライアントは、キャッシュ装置から通知されたデータ精度でよいかどうかの判断を行い(83)、良ければ、キャッシュ装置にその旨を伝える(803)。キャッシュ装置は、肯定応答をクライアントから受け取ると圧縮されたデータをキャッシュから取り出して転送する(804)。

【0011】図8(b)はクライアントの要求がキャッシュサーバの圧縮キャッシュにヒットしたが、当該圧縮によるデータの品質の劣化が許容できない場合の処理の流れを示す図である。まず、クライアント1がキャッシュ装置2に対してデータアクセスを要求する(805)。次にキャッシュ装置はオリジナルデータのキャッシュ（図示略）と圧縮データのキャッシュ（図示略）を調べ、ここで、圧縮キャッシュにのみヒットした場合には、キャッシュ装置はクライアントに対して、デー

(4)

特開平11-219312

5
の精度を応答する(806)。すると、クライアントはデータの品質の精度の判定を行い(84)、データ品質が許容できない否定応答をキャッシュ装置2に送す(807)。すると、キャッシュ装置はデータ供給源5のオリジナルデータに対してアクセス要求を発行し(808)、オリジナルデータ85をクライアントに転送する(809)。

【0012】図8(c)はクライアントからの要求が、キャッシュサーバ2の圧縮キャッシュにもミスした場合の処理の流れを示す図である。まず、クライアント1が10
キャッシュサーバ2に対してデータアクセスを要求する(810)。次にキャッシュサーバは記憶域に保持するオリジナルデータのキャッシュと圧縮データのキャッシュを調べる。ここで、両方のキャッシュにミスすると、キャッシュサーバ2は、データ供給源5に格納されているオリジナルデータ85にアクセス要求を発行する(811)。データ供給源5は、オリジナルデータ85をキャッシュサーバ経由でクライアント1に転送する(812)とともに、キャッシュサーバ2は、転送されてきた20
オリジナルデータ85をキャッシュに登録する。ここで、転送されてきたオリジナルデータのサイズが大きく、オリジナルキャッシュに格納できない場合には、そのデータをシステムの状態に応じて定められたサイズのデータに圧縮して(88)、圧縮キャッシュに登録する(87)。

【0013】

【発明が解決しようとする課題】上記で述べた、圧縮キャッシュを実現する従来技術では、クライアントとキャッシュサーバの間のアクセスインタフェースの整合性を保つために、圧縮したファイルにアクセス要求がヒット30
したときのみクライアントにそのデータ精度でよいかどうかの問い合わせを行うプロトコルを採用していた。しかしながら、このようなプロトコルでは、圧縮ファイルにヒットしたときにクライアントとキャッシュサーバ間の通信回数が2往復必要である。従って、クライアントとサーバ間を接続するネットワークに対する負荷が大きくなってしまおうという問題点があった。

【0014】また、データの精度を認識するプロトコルを実装するためには、クライアント側とサーバ側の両方の通信プロトコルを変更する必要があり、さらに、クライアント側には精度を認識するための判定機構を装備しなければならないという問題点があった。キャッシュサーバの側に対してクライアントの側は膨大であるため、このような変更を行うことは非常に困難であった。

【0015】また、図8では、便宜上クライアント1とデータ供給源5は1台のキャッシュ装置2を介して接続されているように描いているが、実際には、図8(b)に示すように、クライアント1とデータ供給源5の間には複数のキャッシュ装置(201~203)が存在する。従って、圧縮キャッシュの精度がクライアントに拒30

6
絶されるのと同様(901)、キャッシュ装置からの次の中キャッシュ装置に対してデータアクセス要求をする際に、同様のプロトコルによりデータ精度の拒絶が繰り返し行われ(902~904)、データアクセスのレイテンシの著しい低下を招くという問題点があった。

【0016】さらに、従来のデータアクセス方法では、データ供給源からデータをアクセスする場合には必ずオリジナルデータが転送されるため、たとえ、ユーザがそれほど高い品質のデータを求めている場合でも大量のデータがネットワーク上を流れる(905~908)という問題点がある。

【0017】本発明は、大規模データをキャッシングする際に、ユーザの要求やキャッシュ装置の状態に応じて圧縮格納し、要求に応じて圧縮されたキャッシュデータをクライアントに送す方法を提供することを目的とする。また、ネットワーク負荷の軽減を図るために、データ供給源でデータを圧縮して転送するための方法をも提供することを目的とする。さらに、本発明は、上記データの精度をクライアントからのデータアクセス要求時にヘッダ情報として付加する機構を提供することで、キャッシュサーバや情報供給源がユーザの要求にあったデータを効率よく提供するデータアクセス方法を提供することを目的とする。

【0018】

【課題を解決するための手段】上記の目的を達成するため、本発明では、クライアントがデータアクセス要求を発行するときにアクセスするデータIDだけでなくデータの要求精度も同時に付加してアクセスする。オリジナルデータを格納している場合にはデータの要求精度を指定しないようにする。要求精度付きのアクセス要求を受け付けたサーバは、通常のキャッシュ領域以外にも圧縮キャッシュ領域を検索し、圧縮キャッシュ領域にヒットした場合には、要求精度と格納されているデータの圧縮精度とを比較し、それが満たされていれば、圧縮データを返す。圧縮キャッシュにミスした場合や、精度が満たされていない場合には、データ供給源、あるいは、よりデータ供給源に近いキャッシュサーバに対して、クライアントから送られてきた精度付きデータアクセス要求を発行してデータの取得を行う。

【0019】また、データ供給源は、従来、データの要求精度を無視してオリジナルデータを必ず送すように設計されている。しかし、このデータ供給源のサーバに関しても、内部で圧縮データを保持する手段と、精度付きのアクセス要求を解釈して、精度が満たされているかどうかを判断する手段を備えることにより、データ供給源が、精度付きデータアクセス要求に対して圧縮データを応答することができる。これにより、送出すべきデータ量が大幅に削減できるため、データ供給源のデータ送出コストや、ネットワーク負荷を大幅に削減できる。

【0020】本発明によるデータアクセスでは、データ

(5)

特開平11-219312

アクセス時にデータの要求精度を指定するだけで、クライアントが意図するデータを取得することができる。そのため、クライアント側の修正は最小限に抑えられる。さらに、データのアクセス精度を指定しない場合には、オリジナルデータを要求していると解釈することにより、常に正確なデータを得ることを前提としている従来のクライアントアプリケーションを修正する必要は全くない。

【0021】また、データの要求精度は、データアクセス時のヘッダ情報として組み込み、かつ、上記ヘッダ情報はオプションの領域を利用することで、精度付きのデータアクセス要求に対応していない従来のキャッシュサーバはこの情報を容易に無視することが可能であり、従って、本発明の機能を実装した新しいキャッシュサーバと、従来型のキャッシュサーバがネットワーク上で存在していたとしてもうまく機能することが保証される。

【0022】以上の理由により、本発明によるキャッシュ方法や、それを利用することを前提としたクライアントアプリケーションは、従来システムに容易に追加できる。

【0023】

【発明の実施の形態】図1は本発明を実施するキャッシュ装置の概要を示す図である。クライアント1から発行された精度付きデータアクセス要求101は、キャッシュ装置2に送られる。キャッシュ装置2では、アクセス要求のデータ識別子から該当するデータ識別子配列21を検索し、該当データがキャッシュ上にあるかどうかを調べる。キャッシュ装置2には、データ識別子配列21の他に、データを格納するキャッシュ領域23、圧縮属性を格納する圧縮属性配列22が存在する。但し、データそのものに圧縮属性を添付してキャッシュ領域に一緒に格納する場合に、圧縮属性配列22は必要としない。

【0024】キャッシュ装置2では、アクセス要求のあったデータがキャッシュに存在（ヒット）した場合に、前記データの圧縮属性とアクセス要求に記述されているデータの要求精度を比較し、圧縮属性に基づくデータの精度がデータの要求精度を満足しているならば、圧縮されたデータと圧縮属性をクライアント1に返送する。クライアント1では、必要に応じて、受け取った圧縮データを圧縮属性に従って伸張処理を行い、データを表示する。

【0025】ここで、データを圧縮するために用いられる手段としては、映像データには、MPEGなどに代表される任意の動画圧縮技術や、画像データには、JPEGやフラクタル圧縮などに代表される任意の静止画圧縮技術が適用でき、音声に関しては種々の圧縮技術が適用できる。また、精度を落とすことも許されるため、例えば、動画に関しては、1秒当たりのフレームを間引いたり、あるいは、シーン毎の代表画像のみに圧縮するな

と、通常の圧縮とは少し違った形でのデータ圧縮も適用可能である。また、テキストデータの圧縮に関しては、全文の内容からキーワードのみを抽出してキャッシングしたり、あるいは、文章を自動的に要約してキャッシングする技術が適用可能である。さらに、リレーショナルデータのテーブルなどはサンプルデータをキャッシングしておくことなどの種々のデータ圧縮（あるいはデータ量削減）技術が適用できる。

【0026】クライアント1からのアクセス要求がキャッシュ装置2が保持するキャッシュに存在しない（キャッシュミスした）場合には、キャッシュ装置2は、装置に接続されている2次記憶装置4や、ネットワークなどを介して接続されている外部情報システム3等のデータ供給源に格納されているデータをアクセスする。アクセスされたデータは、クライアント1が指定した精度と、キャッシュ装置に設定されたパラメータ等を参照して、圧縮方法が決定され、圧縮手段24によって圧縮されてクライアントに返送される。さらに、圧縮されたデータは、データ識別子、圧縮情報等とともに、キャッシュ装置2に格納される。

【0027】圧縮方法の決定には、クライアントが指定した精度を満足するように、かつ、キャッシュ装置のシステム性能で制限された、利用可能圧縮手法、標準圧縮方法、標準圧縮率、キャッシュ可能データの最大量などを満たすように決定する。これらを満足するような圧縮方法が見つからない場合（例えば、クライアントの要求精度が高いためにキャッシュ可能データの最大量を越える場合など）には、圧縮処理は行われず、情報源から取得したデータを圧縮せずにクライアントに返送する。

【0028】上記の図1で説明した例では、データの圧縮をキャッシュ装置内で行っていた。本発明では、データの圧縮をデータ供給源で行うことももちろん可能である。

【0029】図2はデータ供給源で圧縮処理を行う場合の本発明の実施例である。クライアント1から発行された精度付きデータアクセス要求101は、キャッシュ装置2に送られる。キャッシュ装置2では、アクセス要求のデータ識別子から該当するデータ識別子配列21を検索し、該当データがキャッシュ上にあるかどうかを調べる。クライアント1からのアクセス要求がキャッシュ装置2が保持するキャッシュに存在しない（キャッシュミスした）場合には、キャッシュ装置2は、データ供給源に格納されているデータをアクセスする。ここで、データ供給源が、データ処理装置31とデータ格納装置32で構成されている場合は、キャッシュ装置2はデータ供給源に対して精度付きのデータアクセス要求を発行する。すると、データ供給源が、要求されたデータ精度を満たすように、データ処理装置31内部に備える圧縮手段33によってデータ格納装置32に格納されているデータを圧縮し、キャッシュ装置2に圧縮されたデータと

9

圧縮情報を返送する。

【0030】キャッシュ装置2がデータ供給源に対して精度付きのデータアクセス要求を発行するときには、上記データの精度は、クライアント1からの要求精度を満たし、かつ、キャッシュ装置2のシステム負荷で制限されたキャッシュ可能データの最大量を超えない等の条件を満たすように決定する。なお、キャッシュ装置の制限とクライアント1からの要求精度の両方を満たすことができない場合には、クライアント1からの要求精度をそのままデータ供給源に対して要求する。また、キャッシュ装置2の処理の負荷を削減するためやデータアクセスのレイテンシを削減するために、データ供給源に対して要求するデータ精度はクライアント1と同一としておき、データが供給源から返送されてきたときに、クライアント1にそのデータをそのまま返送するとともに、キャッシュ装置2内に取り込めるかどうかをその時点で判断しても良い。その場合、キャッシュ装置2内に取り込めないと判断された場合には、その圧縮データのキャッシングをあきらめても良いし、また、異なる圧縮をキャッシュ装置2内で行い、キャッシングできる形に再加工してからキャッシングしても良い。

【0031】上記の図2で説明した例では、データの圧縮はデータ供給源に対してキャッシュ装置2からアクセス要求が発行されたときに行われる。従って、ネットワークよりもデータ供給源の処理能力がシステム全体のネックとなっていた場合には、データ取り出しまでに時間がかかる可能性がある。そこで、あらかじめデータ供給源でも頻りにアクセスされる大規模データを圧縮しておく方法について次に述べる。

【0032】図3はデータ供給源であらかじめオリジナルデータとは別に圧縮されたデータを持っている場合に本発明を適用するときの例を示した図である。クライアント1から発行された精度付きデータアクセス要求101は、キャッシュ装置2に送られる。キャッシュ装置2では、アクセス要求のデータ識別子から該当するデータ識別子配列21を検索し、該当データがキャッシュ上にあるかどうかを調べる。クライアント1からのアクセス要求がキャッシュ装置2が保持するキャッシュに存在しない（キャッシュミスした）場合には、キャッシュ装置2は、データ供給源に格納されているデータをアクセスする。この際、前述した方法と同様にキャッシュ装置2はデータ供給源に対して精度付きのデータアクセス要求を発行する。データ供給源は精度付きのアクセス要求を受け取ると、圧縮済みのデータがあるかどうかを判定し、存在した場合には、その圧縮によるデータの精度がアクセス要求精度を満たすかどうかを判定し、満足する場合には、圧縮されたデータをキャッシュ装置2に転送する。

【0033】図4は本発明によるキャッシュ装置の内部構成の実施例を示す図である。クライアントが発行した

(5)

特開平11-219312

10

精度付きのアクセス要求11はキャッシュ装置内でデータ識別子とデータ精度を含む、キャッシュ装置内では、まずデータ識別子情報を用いてデータ識別子配列21のエントリを検索し、関連づけられた圧縮属性及び圧縮データを取り出す。取り出されたデータ識別子は、アクセス要求のデータ識別子との照合処理が行われる（28）。一方、取り出された圧縮属性はデータの要求精度と比較され、要求精度を満たされるかどうかを判断する（27）。この例の場合には、要求精度がJPEG1/100圧縮で、キャッシュの圧縮属性がJPEG1/50圧縮であるので、要求精度を満たす。これらの2つの判定結果はヒットミス判定処理部（28）に送られ、2つとも満たされた場合には、キャッシュヒット信号（29）を生成する。また、取り出された圧縮データは、圧縮属性とともにクライアントに返送される（12）。

【0034】図5は本発明による、キャッシュミスが起きた場合の動作の例を示す図である。クライアントからの要求がキャッシュにミスした場合、そのミス通知（40）を受けたデータアクセス処理手段41がデータ供給源に対してデータアクセス要求を発行する。データ供給源から得られたデータの属性情報、クライアントの要求精度、キャッシュ装置の特性を示すシステム情報を考慮し精度決定処理手段42が圧縮方法を決定し、圧縮手段4よりデータ圧縮を行う。圧縮されたデータは、クライアントに返送され、さらに、データ識別子、圧縮属性とともにキャッシュ装置2内のデータ識別子配列、圧縮属性配列、キャッシュ領域にそれぞれ格納される。

【0035】本実施例では、キャッシュミスした場合の圧縮処理を、図1で説明したキャッシュ装置内で実行している実施例に対応する詳細図を示しているが、図2及び図3で示した実施例に対応する詳細図に関しても全く同様に記述できる。その場合には、圧縮処理手段24と精度決定処理手段42がデータ供給源側に備わっており、データアクセス手段から発行されるデータアクセス要求は精度付きアクセス要求となる。

【0036】図6は本発明によるデータアクセス手順を示した図である。クライアント1はデータ供給源3にあるデータをキャッシュ装置2経由でアクセスする。ここで、クライアント1はデータ供給源3にあるオリジナルデータでなく、精度を一定以上に保ったデータを要求しているとする。データの精度としては、例えば、画質、書質、文章の要約度、DBのサンプリング率、統計情報の精度などがある。キャッシュ装置2は、オリジナルデータのキャッシュ（オリジナルキャッシュ）B1と、オリジナルデータを圧縮したキャッシュ（圧縮キャッシュ）B2を保持する。圧縮キャッシュに格納されるデータは、主にサイズが大きいデータで、かつ、データの精度を落とすことがある程度許容されるデータを格納する。データの精度をどの程度落として格納するかは、デ

(7)

特開平11-219312

11

ータのアクセス履歴を保管することにより判定できる。すなわち、クライアントから発行されるデータ識別子と要求精度の履歴を取ることに、どのデータの精度を落とすことが可能かどうかを判断することが可能となる。もちろん、例えば、50KB以上の静止画像は1/100のJPEG圧縮といったように、データサイズやデータ種別により一元的に圧縮するか否かを決定する方法を取ることも可能である。

【0037】図6(a)はクライアントからの精度付きアクセス要求がキャッシュ装置の圧縮キャッシュにヒットした場合のデータアクセス手順を示す図である。クライアントから発行された精度付きデータアクセス要求801はキャッシュ装置に送られる。キャッシュ装置は精度付きのアクセス要求であるので、オリジナルキャッシュならびに圧縮キャッシュの両方にアクセスを行う。このとき、オリジナルキャッシュにのみヒットした場合には、そのデータをクライアントに直接送達する。一方、図に示すように圧縮キャッシュにのみヒットした場合にはデータの圧縮方法が要求精度を満たすか否かの判定を行い(83)、精度を満たしていた場合には圧縮データを圧縮属性とともにクライアントに転送する。

【0038】なお、圧縮キャッシュとオリジナルキャッシュの両方にヒットし、かつ、データの要求精度が増たされている場合には、データ転送料の削減のため、原則として圧縮キャッシュのデータを優先的にクライアントに返す。但し、キャッシュ装置とクライアントの間のデータ転送のためのコストが小さい場合には、オリジナルデータをクライアントに転送してもよい。

【0039】図6(b)に示した従来方式では、圧縮キャッシュにヒットした場合、データの精度の追跡手順が必要であったが、データの要求精度そのものをアクセス要求と一緒にキャッシュ装置に送り込むことで、データの精度の追跡手順を省略化することができる。また、データ精度の判定処理をキャッシュ装置内で行うため、クライアントシステムに大幅な変更を加える必要がない。さらに、データアクセスの要求精度がキャッシュ装置に毎回伝えられるため、上記要求精度の履歴をキャッシュ装置内で保持することにより、適切な精度でデータを圧縮してキャッシュすることが可能となる。

【0040】図6(b)はクライアントからのアクセス要求が、キャッシュ装置のオリジナルキャッシュ、圧縮キャッシュの両方にミスした場合のアクセス手順を示す図である。クライアントから発行された精度付きのアクセス要求が、キャッシュ装置のオリジナルキャッシュ及び圧縮キャッシュにミスした場合(圧縮キャッシュにミスする場合には、圧縮キャッシュのデータにはヒットしたがデータ精度が要求精度を満たさなかった場合も含む)には、クライアントからの精度付きアクセス要求がデータ供給源5に対して送られる(803)。データ供給源では、受け取った精度付きアクセス要求を満たすデ

12

ータをデータ供給源が保持するオリジナルデータならびに圧縮済みデータから検索する。ここで、圧縮済みのデータにアクセス要求がヒットし、アクセス要求精度が満たされていると判定される(804)と、上記圧縮済みデータが圧縮属性とともにキャッシュ装置に送られる(804)。キャッシュ装置は受け取った圧縮済みデータをクライアントに転送する(802)とともに、上記データを圧縮キャッシュに登録する(805)。

【0041】本発明では、キャッシュ装置にミスした場合でも、精度付きのアクセス要求をデータ供給源に対して発行するため、データ供給源がアクセス要求精度を満たす圧縮データを持っている場合には、圧縮データのみを転送することで、クライアントのアクセス要求を処理することができる。本方式により、オリジナルデータを転送する必要がないため、ネットワーク負荷の軽減ができ、かつ、データ供給源のデータ送出手の負荷も軽減される。データ供給源の圧縮データの管理や、精度付きアクセス要求に対するヒット判定方法は、キャッシュ装置と同様の方法で実現できる。

【0042】なお、キャッシュ装置からデータ供給源に対して発行される精度付きアクセス要求は、クライアントからの精度付きアクセス要求そのまま発行してもよいし、あるいは、キャッシュ装置内に格納されているデータアクセス履歴やキャッシュ装置の制限などにより、クライアントからのデータ要求精度が増たされる範囲で修正を加えた上で発行してもよい。

【0043】図7はクライアントのアクセス要求が、キャッシュ装置にミスし、さらに、データ供給源の圧縮データにもミス(あるいは、データ供給源に圧縮データが最初から存在しない)した場合のデータアクセス手順を示す図である。この場合、データ供給源に格納されているオリジナルデータに対してアクセスする必要がある。

【0044】図7(a)はデータ供給源から、オリジナルデータをキャッシュ装置に転送する場合の例を示す。クライアントから発行された精度付きアクセス要求701はキャッシュ装置に送られ、そこで、キャッシュのヒット判定が行われる。ここで、キャッシュにミスすると、データ供給源に対して精度付きアクセス要求703が発行される。データ供給源では、受け取ったアクセス要求に基づきデータを検索するが、ここで、適切な圧縮データがない場合には、オリジナルデータをそのままキャッシュ装置に転送する(704)。キャッシュ装置では、受け取ったデータをクライアントのアクセス要求精度に応じて圧縮し(71)、その結果をクライアントに返送する(702)。さらにキャッシュ装置は、圧縮されたデータをキャッシュにも登録する(72)。

【0045】この例では、データ供給源からキャッシュ装置まではオリジナルデータを転送している。そのため、データ供給源にはオリジナルデータの転送負荷のみがかかり、圧縮処理の負荷はキャッシュ装置にかかる。

(8)

特開平11-219312

13

そのため、データ供給源の処理性能が低い場合に特に有効である。

【0046】しかしながら、データ供給源の処理能力が十分大きい場合には、図7(b)に示すようにデータ圧縮処理をデータ供給源側で行う方法が良い。この場合、キャッシュ装置から受け取った精度付きデータアクセス要求703に応じてオリジナルデータを圧縮し(703)、その結果を圧縮装置とともにキャッシュ装置に転送する(705)。このようにすることで、データ供給源の負荷は高くなるが、ネットワークを流れるデータの量が大幅に削減でき、また、キャッシュ装置の負荷も軽減できるという利点がある。さらに、データ供給源側で圧縮したデータを圧縮データ格納装置に保存しておくことにより、次のデータ供給源に対する精度付きデータアクセス要求時に上記データを再利用できるという利点がある。

【0047】図7のいずれの方法を採るかは、サーバの負荷と、ネットワークの負荷のトレードオフにより決定できる。いずれにしても、キャッシュ装置側としては、精度付きのアクセス要求をデータ供給源に対して発行し、受け取ったデータが圧縮済みであれば、そのデータをクライアント側にそのまま渡し、キャッシュにも登録すればよいし、受け取ったデータがオリジナルデータであれば、圧縮してから同様の手続きをとればよいので、これらの両方の方式は混在可能である。

【0048】なお、以上の説明の例では、キャッシュ装置から直接データ供給源にアクセスするように説明したが、もちろん、データ供給源との間に同様の働きをするキャッシュ装置が介在してもかまわない。

【0049】図8は、クライアントとデータ供給源との間に複数のキャッシュ装置(201~103)が存在する場合を示す図である。図9(a)は従来方式によるデータのアクセス手順である。この場合には複数のキャッシュ装置間でデータアクセス要求・データ精度の確認・データ精度の伝達のプロトコルが繰り返し実行され(901~904)データアクセス性能が劣化する。従来方式では、最初のキャッシュ装置201のキャッシュにミスするとオリジナルデータをアクセスするため、後続のキャッシュ装置(202~203)に圧縮されたキャッシュデータがあり、かつ、その精度がクライアントの要求を満たしている場合でも上記圧縮データを利用することはない。

【0050】図9(b)は本発明による圧縮キャッシュのアクセス方法を適用した場合のアクセス手順の概要を示す図である。本発明では、クライアントからのアクセス要求にデータの要求精度が記載されているため、圧縮キャッシュにデータが存在したときにクライアントにデータ精度の確認を行う必要がない(911~914)。また、本方式では、データ精度を含むアクセス要求を発行するため、データ供給源5が圧縮したデータを保持して

14

いるか、あるいはデータ供給源5がアクセス要求時に圧縮処理を施せば、上記圧縮したデータをクライアントに返送することが可能である(915~918)。このため、ネットワーク中を流れるデータを従来方式に比べ大幅に削減することができる。

【0051】さらに、キャッシュ装置にミスした場合には、データの要求精度付きのアクセス要求を後続のキャッシュ装置(202~203)あるいはデータ供給源5に発行するので、最初のキャッシュ装置にミスした場合での途中のキャッシュ装置の圧縮データがデータの要求精度を満たしているならば、その圧縮データを利用してクライアントのアクセス要求に応えることができる。

【0052】図9(c)はキャッシュ装置201にミスしたアクセス要求がキャッシュ装置202の圧縮キャッシュにヒットした場合の処理の流れを示す図である。キャッシュ装置201にミスしたアクセス要求(921)は、データ精度付きのアクセス要求のままキャッシュ装置202に送られる(922)。したがって、キャッシュ装置202が上記要求精度を満たすデータをキャッシュとして保持していれば、そのデータをクライアントへ返送する(923~924)ことができる。

【0053】

【発明の効果】情報システムにおけるデータオブジェクトのキャッシュならびにアクセス方法として、サイズの大いデータを圧縮してキャッシュ装置に格納するとともに、データアクセス要求時にデータの要求精度もキャッシュ装置やデータ供給源に通知することにより、圧縮されたデータをキャッシュとして有効に利用することが可能となり、データアクセス時間の削減、ネットワーク負荷の削減、サーバの負荷の削減等を行うことが可能となる。

【図面の簡単な説明】

【図1】本発明の一実施例における、データキャッシュ及びアクセス方法の概要を示す説明図。

【図2】データ供給源でデータ圧縮処理を行うときの処理の流れを示す説明図。

【図3】データ供給源であらかじめ圧縮データを保持している場合の処理の流れを示す説明図。

【図4】キャッシュ装置の圧縮キャッシュにヒットしたときの処理の流れを示す説明図。

【図5】キャッシュ装置の圧縮キャッシュにミスしたときの処理の流れを示す説明図。

【図6】圧縮キャッシュあるいは圧縮データにヒットしたときのデータアクセス手順の説明図。

【図7】圧縮キャッシュ及び圧縮データにミスしたときのデータアクセス手順の説明図。

【図8】従来の圧縮キャッシュを用いたデータアクセス手順の説明図。

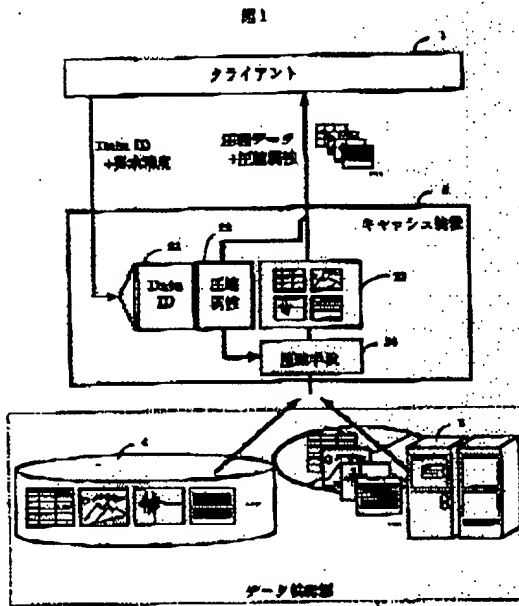
【図9】キャッシュ装置が複数接続されているときのアクセス手順の説明図。

15

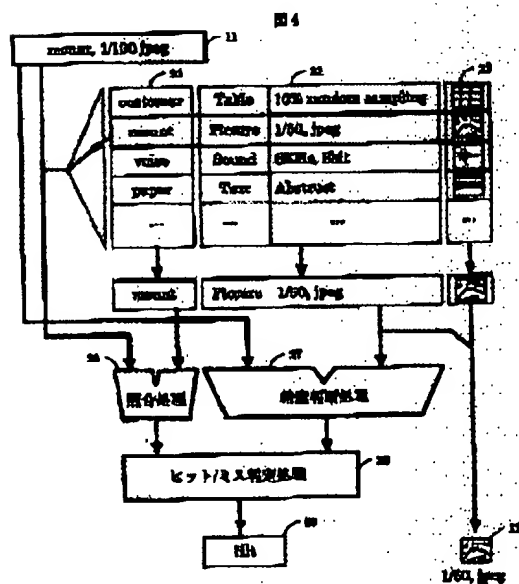
【符号の説明】

1…クライアント、2…キャッシュ装置、3…外部情報システム、4…二次記憶装置、5…データ供給線、6…*

【図1】



【図4】



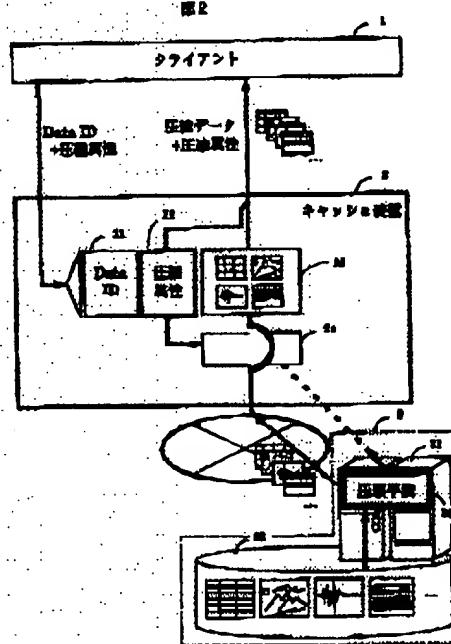
(9)

特開平11-219312

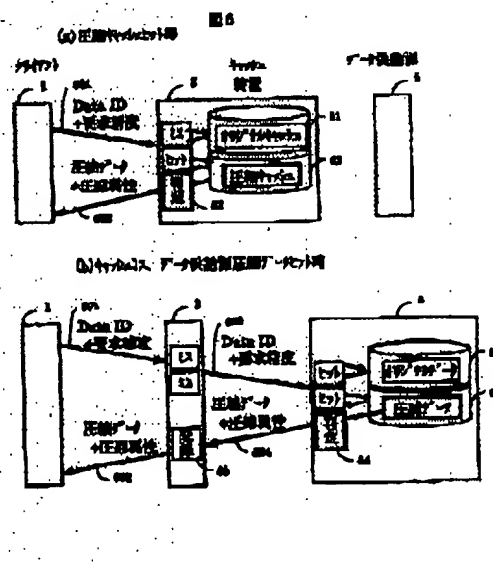
16

*ネットワーク、21…データ識別子配列、22…圧縮/展開性配列、23…キャッシュ領域。

【図2】



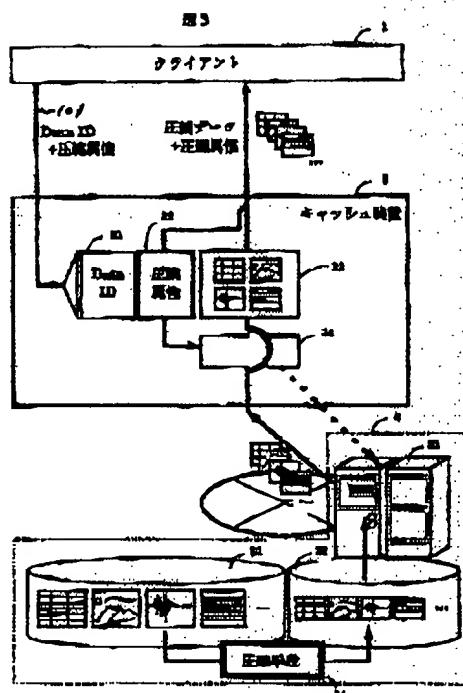
【図6】



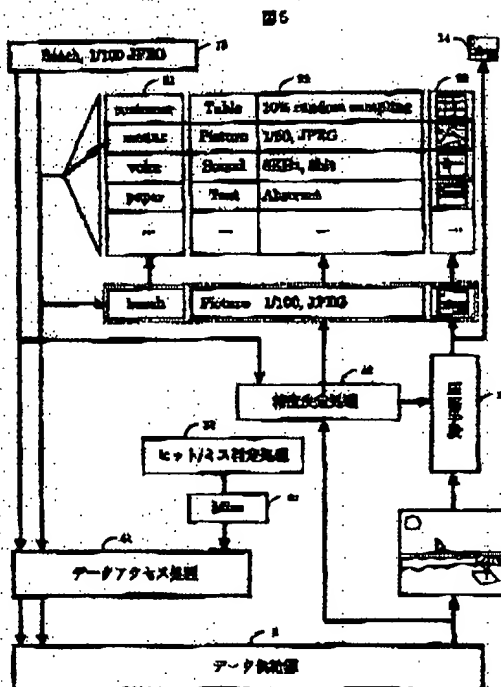
(10)

特開平11-219312

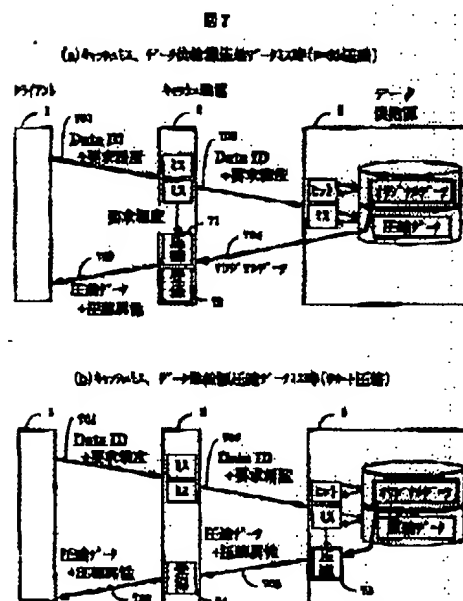
【図3】



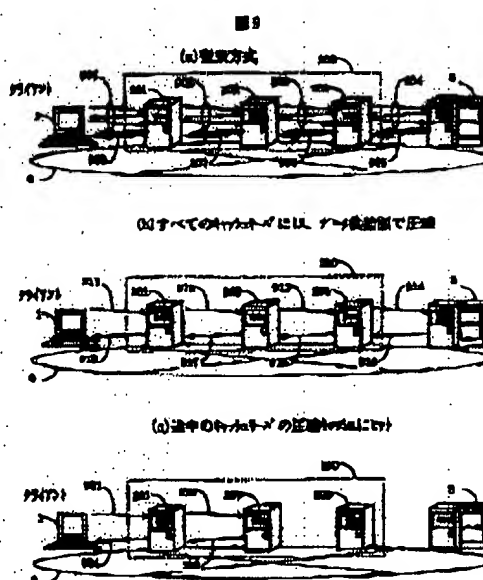
【図5】



【図7】



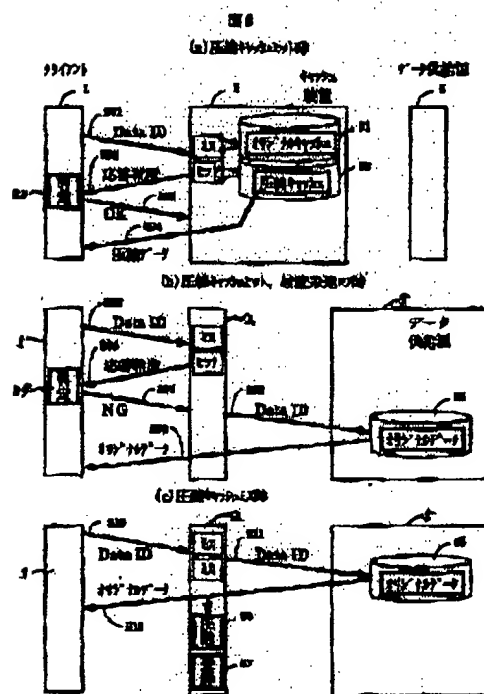
【図9】



(11)

特開平11-219312

【図8】



フロントページの続き

(72)発明者 猪原 茂和

東京都国分寺市東厩ケ腰一丁目280番地

株式会社日立製作所中央研究所内